Modified likelihood ratio test for homogeneity in a mixture of von Mises distributions

Yuejiao Fu

Department of Mathematics and Statistics, York University

Toronto, ON, M3J 1P3, Canada

yuejiao@mathstat.yorku.ca

Jiahua Chen and Pengfei Li

Department of Statistics and Actuarial Science, University of Waterloo

Waterloo, ON, N2L 3G1, Canada

jhchen@uwaterloo.ca   p4li@uwaterloo.ca

SUMMARY. Directional data are often collected in applications such as in genetics, geology and astrolomy. One particular statistical problem of interest is whether the data are from a mixture of two von Mises distributions or a single von Mises distribution. Motivating examples including a DNA microarray experiment where it is suggested that a proportion of circadian genes have systematically different phase/peak expressions in two different tissues. We study the use of modified likelihood ratio test (MLRT) to this class of problems. The MLRT statistic is shown to have a simple χ_1^2 null limiting distribution. The result is extended to general parametric kernels. The simulation study gives additional insight into the finite-sample performance of the test. For illustration, two real data examples are also given.

KEY WORDS: Asymptotic distribution; Circadian gene; Mixture of von Mises distributions, Modified likelihood.

1. Introduction

Directional data are often collected in applications such as in genetics, geology and astrolomy. One particular statistical problem of interest is whether the data are from a mixture of two von Mises distributions or a single von Mises distribution. Motivating examples including a DNA microarray experiment where it is suggested that a proportion of circadian genes have systematically different phase/peak expressions in two different tissues. An interesting biological problem is to compare the activation times (phase angles) of a set of circadian-related genes in two or more tissues. Let ϕ_i^x and ϕ_i^y denote the phase angles of a circadian-related gene i, $i = 1, 2, \ldots, n$, in the two tissues x and y, where $|\phi_i^x| \leq \pi$, $|\phi_i^y| \leq \pi$. The distribution of the angular difference $\theta_i = \phi_i^x - \phi_i^y$ may be modeled as a two-component von Mises mixture. One component corresponds to the case where a subset of the n genes have the same phase angle in the two tissues and the other corresponds to a set of genes having a discrepancy in phase angle between the two tissues. The problem of interest is to make statistical inference on the existence of heterogeneity in phase angle difference.

The likelihood ratio test (LRT) is the most extensively used method for parametric hypothesis testing problems. It is well known that under the standard regularity conditions, the LRT has a chi-squared null limiting distribution. Due to the non-regularity of mixture models, the usual LRT often has a complex limiting distribution (Dacunha-Castelle and Gassiat 1999; Liu and Shao, 2003) and therefore loses much of its appeal in statistical inference. The modified likelihood ratio test (MLRT), proposed by Chen (1998), Chen et al. (2001, 2004) and Chen & Kalbfleisch (2005), provides a nice solution to this problem by simply adding a penalty term to the log-likelihood function. The limiting distribution of the MLRT statistic is chi-squared or a mixture of chi-squared distributions for a large variety of mixture models. The modified likelihood method has the advantage of giving a natural and quite general approach to testing problems in finite mixture models.

In this paper, we investigate the use of the MLRT to test homogeneity in a mixture of von Mises distributions. The ordinary maximum likelihood estimator (MLE) of the concentration parameter in the von Mises mixture model is shown to be consistent. The asymptotic null distribution of the regular LRT statistic is proven to be a squared supremum of truncated Gaussian process. Based on this result, we show that the MLRT statistic has a very simple χ_1^2 limiting distribution which can be easily applied. We also extend the result to a mixture model with a general parametric kernel. There are a variety of real application examples in the literature which are special cases of this formulation. In particular, the results are applied to circular data discussed in Liu et al. (2006) in genetic research, and a directional data of dinosaur bones in geological investigation(Grimshaw, Whiting and Morris, 2001) .

The remainder of this paper is organized as follows. We layout the problem in Section 2. The main results are presented in Section 3 and the extension to general parametric kernels is given in Section 4. In Section 5, we conduct a simulation study to evaluate the finite-sample performance of the MLRT. Further, we apply the MLRT to two real data examples. Finally, we conclude with some discussion in Section 6. The mathematical details and proofs of the theorems are deferred to the Appendix.

2. Problem Setup

Suppose we observe a circular (angular) random sample $\theta_1, \ldots, \theta_n$ from a mixture population $(1-\alpha)M(0,\kappa) + \alpha M(\mu,\kappa)$, where $0 \le \alpha \le 1$, $|\mu| \le \pi$ and $\kappa \ge 0$. Here $M(\mu,\kappa)$ denotes the von Mises distribution with mean direction μ and concentration parameter

κ. The special feature of this mixture population is that the mean direction of one component is known to be zero and both of the components have the same concentration parameter.

The von Mises distribution was first introduced by von Mises (1918). As a circular analog of the Normal distribution on the real line, it is also called the circular Normal distribution. Similar to the Normal distribution for linear data, it is the most commonly used distribution for circular data, see Mardia & Jupp (2000) for general properties of the von Mises distribution. The probability density function (pdf) of the von Mises distribution is

$$f(\theta; \mu, \kappa) = \frac{1}{2\pi I_0(\kappa)} \exp\{\kappa \cos(\theta - \mu)\}, \quad |\theta| \le \pi, \tag{1}$$

where $|\mu| \le \pi$ and $\kappa \ge 0$. The pdf of the von Mises distribution (1) is unimodal and symmetric about $\theta = \mu$. When $\kappa = 0$, the von Mises distribution becomes a uniform circular distribution and when $\kappa = \infty$ a point distribution. The function $I_0(\kappa)$ is defined by

$$I_0(\kappa) = \frac{1}{2\pi} \int_0^{2\pi} \exp(\kappa \cos\theta) d\theta,$$

which is the normalizing constant and known as the modified Bessel function of the first kind and order zero. In general, the modified Bessel function I_p of the first kind and order p (sometimes also called Bessel function of purely imaginary argument) can be defined by

$$I_p(\kappa) = \frac{1}{2\pi} \int_0^{2\pi} \cos(p\theta) \exp(\kappa \cos\theta) d\theta.$$

Usually, we use $A(\kappa)$ to denote the ratio of the two modified Bessel functions,

$$A(\kappa) = \frac{I_1(\kappa)}{I_0(\kappa)}. \tag{2}$$

Properties of these functions can be found in Abramowitz & Stegun (1965). In this paper, our aim is to investigate statistical methods for testing

$$H_0 : M(0, \kappa) \text{ versus } H_a : (1-\alpha)M(0, \kappa) + \alpha M(\mu, \kappa). \tag{3}$$

The focus of this paper is on the asymptotic properties of likelihood-based testing procedures.

3. Main Results

Let $\theta_1, \ldots, \theta_n$ be a circular random sample from the mixture population $(1-\alpha)M(0, \kappa) + \alpha M(\mu, \kappa)$. The log-likelihood function can be expressed as

$$l_n(\alpha, \mu, \kappa) = -n \log I_0(\kappa) + \sum \log[(1-\alpha)\exp(\kappa \cos\theta_i) + \alpha \exp\{\kappa \cos(\theta_i - \mu)\}]. \tag{4}$$

Let $\hat{\kappa}_0$ be the MLE under the null hypothesis and let $\hat{\alpha}$, $\hat{\mu}$, and $\hat{\kappa}$ be the MLEs under the full model. Some statistical properties of $\hat{\alpha}$, $\hat{\mu}$, and $\hat{\kappa}$ are investigated in the subsequent sections.

3.1 Large sample behavior of the MLEs

We first prove that $\hat{\kappa}$ is bounded above from infinity in probability asymptotically even if the true distribution is a non-mixture model $M(0, \kappa_0)$.

LEMMA 1 *Assume that the distribution of the random sample $\theta_1, \ldots, \theta_n$ is given by $M(0, \kappa_0)$ for some $\kappa_0 > 0$. Let $\hat{\kappa}$ be the MLE of κ under the full model $(1-\alpha)M(0, \kappa) + \alpha M(\mu, \kappa)$. Then there exists a constant $0 < \Delta < \infty$ such that*

$$\lim_{n\to\infty} P(\hat{\kappa} \leq \Delta) = 1.$$

PROOF. Rewrite the log-likelihood function (4) as

$$l_n(\alpha, \mu, \kappa) = -n \log I_0(\kappa) + \kappa \sum \cos\theta_i + \sum \log\{1 + \alpha[\exp\{\kappa\cos(\theta_i - \mu) - \kappa\cos\theta_i\} - 1]\}.$$

Note that $1+\alpha[\exp\{\kappa\cos(\theta-\mu)-\kappa\cos\theta\}-1]\leq\exp\{\kappa[\cos(\theta-\mu)-\cos\theta]^+\}$, where $[\cdot]^+$ denotes the positive part of the argument. Thus we have

$$l_n(\alpha,\mu,\kappa)\leq -n\log I_0(\kappa)+\kappa\sum\cos\theta_i+\kappa\sum[\cos(\theta_i-\mu)-\cos\theta_i]^+.$$

By (A.4) in Mardia & Jupp (2000, p. 349), $I_0(\kappa)=(2\pi\kappa)^{-1/2}\exp(\kappa)\{1+o(1)\}$, as $\kappa\to\infty$. Hence, for large κ,

$$\begin{aligned}
l_n(\alpha,\mu,\kappa) &\leq -n\kappa+\frac{n}{2}\log(2\pi\kappa)+\kappa\sum\cos\theta_i+\kappa\sum[\cos(\theta_i-\mu)-\cos\theta_i]^+ + o_p(1)\\
&= -\kappa\sum\left[1-\max\{\cos\theta_i,\cos(\theta_i-\mu)\}\right]+\frac{n}{2}\log(2\pi\kappa)+o_p(1).
\end{aligned}$$

Let $S(\mu,\kappa_0)=E\left[1-\max\{\cos\theta,\cos(\theta-\mu)\}\right]$ where κ_0 is the true value of the parameter κ. By the uniform strong law of large numbers (see Rubin, 1956),

$$n^{-1}\sum\left[1-\max\{\cos\theta_i,\cos(\theta_i-\mu)\}\right]\to S(\mu,\kappa_0),$$

almost surely and uniformly in $|\mu|\leq\pi$. For any $|\theta|\leq\pi$, we have the inequality $1-\max\{\cos\theta,\cos(\theta-\mu)\}\geq 0$, where the equality holds only if $\theta=0$ or μ, which has zero probability to occur for any given $0<\kappa_0<\infty$. Therefore, under the null distribution $M(0,\kappa_0)$ with $\kappa_0>0$, $S(\mu,\kappa_0)$ is continuous and positive, for all the values of μ. Thus, $q=\min_{|\mu|\leq\pi}S(\mu,\kappa_0)>0$.

Then with probability approaching one uniformly in α, μ, and κ,

$$l_n(\alpha,\mu,\kappa)\leq -n\{q\kappa-\log(2\pi\kappa)/2\}+o_p(1).$$

Clearly, there exists a $\Delta>0$ such that when $\kappa>\Delta$, we have $q\kappa-\log(2\pi\kappa)/2>0$. Note that $l_n(0,0,0)=0$. The function $l_n(\alpha,\mu,\kappa)-l_n(0,0,0)<0$ in probability when $\kappa>\Delta$. This shows that $\lim P(\hat{\kappa}>\Delta)=0$ for some constant Δ. $\square$

As a consequence of Lemma 1, the parameter space under consideration can be reduced to a compact one for theoretical derivations. Fraser et al. (1981) and Holzmann et al. (2004) proved the identifiability and strong identifiability of finite mixtures of the von Mises distributions. With identifiability, Lemma 1 implies the consistency of the MLEs.

LEMMA 2 *Assume that the distribution of the random sample* $\theta_1, \ldots, \theta_n$ *is given by* $M(0, \kappa_0)$. *Let* $\hat{\alpha}$, $\hat{\mu}$, *and* $\hat{\kappa}$ *be the MLEs of* α, μ, *and* κ *under the full model* $(1-\alpha)M(0,\kappa) + \alpha M(\mu, \kappa)$. *Then* $\hat{\alpha}\hat{\mu} \to 0$ *and* $\hat{\kappa} \to \kappa_0$, *in probability.*

PROOF. The proof is straightforward and follows that of Chen & Chen (2003). □

3.2 Asymptotic distributions of LRT and MLRT

We now study the asymptotic distributions of the LRT and the MLRT. The main results are given in the following two theorems and the proofs are left in the Appendix.

THEOREM 1 *Let* $\theta_1, \ldots, \theta_n$ *be a random sample from the mixture population* $(1-\alpha)M(0,\kappa) + \alpha M(\mu,\kappa)$, *where* $0 \le \alpha \le 1$, $|\mu| \le \pi$ *and* $\kappa \ge 0$. *Let* R_n *be (twice) the log-likelihood ratio test statistic for testing* $H_0 : \alpha = 0$ *or* $\mu = 0$. *Then under the null distribution* $M(0, \kappa_0)$, *as* $n \to \infty$,

$$R_n \to \sup_{|\mu| \le \pi} \{\zeta^+(\mu)\}^2,$$

where $\zeta(\mu)$, $|\mu| \le \pi$, *is a Gaussian process with mean 0, variance 1 and autocorrelation* $\rho(s,t)$ *which is given by*

$$\rho(s,t) = sgn(st) \frac{g(s,t)}{\{g(s,s)g(t,t)\}^{\frac{1}{2}}}, \qquad \text{for } s,t \neq 0, \tag{5}$$

where

$$g(s,t) = \frac{1}{st}\Big[\frac{I_0[\kappa_0\{(\cos s + \cos t - 1)^2 + (\sin s + \sin t)^2\}^{\frac{1}{2}}]}{I_0(\kappa_0)} - 1 - \frac{A^2(\kappa_0)(\cos s - 1)(\cos t - 1)}{1 - A(\kappa_0)/\kappa_0 - A^2(\kappa_0)}\Big].$$

It is noteworthy that the asymptotic distribution of the LRT statistic R_n in the current von Mises mixture model is different from that of the LRT statistic in normal mixture discussed in Chen and Chen (2003). Normal mixtures with unknown variance is not strongly identifiable in the sense that the second derivative of the density with respect to the mean is equal to the first derivative of the density with respect to the variance, which is not the case for the von Mises mixture models. The result on the asymptotic distribution of the LRT provides much insight to the nature of the problem. In order to use this result for the purpose of inference, we need to calculate quantiles of the supremum of the Gaussian process. This is still an open problem in the literature in general (Adler, 1990). Instead, we use the MLRT to address the aforementioned testing problem (). For $0 < \alpha \leq 1,\ |\mu| \leq \pi$, we define the modified log-likelihood function as

$$pl_n(\alpha, \mu, \kappa) = l_n(\alpha, \mu, \kappa) + C \log(\alpha)$$

with $C > 0$ being a specified constant which determines the level of modification. We often take $C = 1$, which has been found to be satisfactory for the data with multinomial component distributions; see Chen (1998). For other mixture models, the appropriate choice of C depends on the size of the parameter space, see Chen et al. (2001) and Zhu & Zhang (2004). The modified log-likelihood ratio statistic is defined by

$$M_n = 2\{pl_n(\hat{\alpha}^*, \hat{\mu}^*, \hat{\kappa}^*) - pl_n(1, 0, \hat{\kappa}_0^*)\},$$

where $(\hat{\alpha}^*, \hat{\mu}^*, \hat{\kappa}^*)$ maximizes $pl_n(\alpha, \mu, \kappa)$ over the region $0 < \alpha \leq 1,\ |\mu| \leq \pi,\ k \geq 0$, and $\hat{\kappa}_0^*$ maximizes $pl_n(1, 0, \kappa)$ which is the modified log-likelihood function under the null hypothesis. The following theorem gives the asymptotic null distribution of M_n.

THEOREM 2 *Let* $\theta_1, \ldots, \theta_n$ *be a random sample from the mixture population* $(1-\alpha)M(0, \kappa) + \alpha M(\mu, \kappa)$, *where* $0 \leq \alpha \leq 1$, $|\mu| \leq \pi$ *and* $\kappa \geq 0$. *Let* M_n *be (twice)* *the MLRT statistic for testing $H_0 : \alpha = 0$ or $\mu = 0$. Then under the null distribution $M(0, \kappa_0)$, the limiting distribution of M_n is χ_1^2.*

The MLRT statistic is asymptotic pivotal and has a very simple limiting distribution under the null hypothesis. It is particularly easy to use in practice. The simulation results in Section 5 also show that χ_1^2 provides a good approximation to the finite sample distribution. The precision is not sensitive to the choice of the level of modification C. Generally, a larger value for C leads to a faster convergence to the asymptotic distribution, but to a lower power of the test.

4. Extension to General Parametric Kernels

The result in Section 3.2 can be extended to a mixture model with a general parametric kernel. Let $\theta_1, \ldots, \theta_n$ be a random sample of size n from a two-component mixture population with the mixture density

$$f(\theta; \alpha, \mu, \kappa) = (1 - \alpha) f(\theta; \mu_0, \kappa) + \alpha f(\theta; \mu, \kappa), \tag{6}$$

where μ_0 is known, $0 \leq \alpha \leq 1$, $\mu \in \mathbf{T}$ and $\kappa \in \mathbf{B}$. Note that the component density $f(\theta; \mu, \kappa)$ belongs to a general parametric family of distributions and two mixture components have a common unknown structural parameter κ. Define

$$U_i(\kappa) = \frac{1}{\kappa - \kappa_0} \Big\{ \frac{f(\theta_i; \mu_0, \kappa)}{f(\theta_i; \mu_0, \kappa_0)} - 1 \Big\}, \; Y_i(\mu, \kappa) = \frac{1}{\mu} \Big\{ \frac{f(\theta_i; \mu, \kappa)}{f(\theta_i; \mu_0, \kappa_0)} - \frac{f(\theta_i; \mu_0, \kappa)}{f(\theta_i; \mu_0, \kappa_0)} \Big\},$$

and $U_i(\kappa_0)$ and $Y_i(\mu_0, \kappa)$ be their continuity limits. For convenience of notation, we put $Y_i(\mu) = Y_i(\mu, \kappa_0)$, $Y_i = Y_i(\mu_0)$, and $U_i = U_i(\kappa_0)$. We have the following regularity conditions on the kernel function $f(\theta; \mu, \kappa)$:

- **A1.** *Compact parameter space.* Both $\mathbf{T}$ and $\mathbf{B}$ are compact subsets of $\mathbf{R}$, and κ_0 is an interior point of $\mathbf{B}$.

A2. *Wald's integrability conditions.* The kernel function $f(\theta;\mu,\kappa)$ satisfies Wald's integrability conditions for consistency of the maximum likelihood estimate.

A3. *Smoothness.* $f(\theta;\mu,\kappa)$ is twice continuously differentiable with respect to μ and κ.

A4. *Identifiability.* The mixing distribution is identifiable.

A5. *Positive definiteness.* The covariance matrix of U_i and $Y_i(\mu)$ is positive definite for all $\mu \in \mathbf{T}$. If κ is known to be κ_0, we only need that $Var\{Y_i(\mu)\} > 0$ for all $\mu \in \mathbf{T}$.

A6. *Uniform strong law of large numbers.* There exists integrable function g such that $|U_i(\kappa)|^3 \leq g(\theta_i)$ and $|Y_i(\mu,\kappa)|^3 \leq g(\theta_i)$ for $\mu \in \mathbf{T}$ and $\kappa \in \mathbf{B}$.

A7. *Tightness.* The processes

$$
\begin{aligned}
&n^{-1/2}\sum\{U_i(\kappa) - U_i\}/(\kappa - \kappa_0),\\
&n^{-1/2}\sum\{Y_i(\mu) - Y_i(\mu_0)\}/(\mu - \mu_0),\\
&n^{-1/2}\sum\{Y_i(\mu,\kappa) - Y_i(\mu)\}/(\kappa - \kappa_0)
\end{aligned}
$$

are tight for $\mu \in \mathbf{T}$ and $\kappa \in \mathbf{B}$.

Note that these regularity conditions are satisfied for both normal and von Mises kernels. Using the similar treatment for the von Mises mixture model, we can show that under conditions $A1 - A7$, the null asymptotic distribution of the LRT statistic is the supremum of a Gaussian process.

We can also consider the MLRT with a general parametric kernel. The modified likelihood function is

$$
pl_n(\alpha,\mu,\kappa) = \sum_{i=1}^{n} \log\{(1-\alpha)f(\theta_i;\mu_0,\kappa) + \alpha f(\theta_i;\mu,\kappa)\} + C\log(\alpha). \tag{7}
$$

Let $(\hat{\alpha}^*, \hat{\mu}^*, \hat{\kappa}^*)$ maximize $pl_n(\alpha, \mu, \kappa)$ over the region $0 < \alpha \leq 1$, $\mu \in \mathbf{T}$ and $\kappa \in \mathbf{B}$. And let $\hat{\kappa}_0^*$ maximizes $pl_n(1, \mu_0, \kappa)$ which is the modified log-likelihood function over the region $\kappa \in \mathbf{B}$. Then the modified likelihood ratio test is to reject the null hypothesis H_0 if

$$M_n = 2\{pl_n(\hat{\alpha}^*, \hat{\mu}^*, \hat{\kappa}^*) - pl_n(1, \mu_0, \hat{\kappa}_0^*)\},$$

is large enough. The following theorem gives the null limiting distribution of M_n. The proof is similar to that of Theorem 2 and therefore omitted.

THEOREM 3 *Let* $\theta_1, \ldots, \theta_n$ *be a random sample from the mixture population* $(1-\alpha)f(\theta; \mu_0, \kappa) + \alpha f(\theta; \mu, \kappa)$, *where* $0 \leq \alpha \leq 1$, $\mu \in \mathbf{T}$, μ_0 *is an interior point of* $\mathbf{T}$ *and* $\kappa \in \mathbf{B}$. *Let* M_n *be (twice) the MLRT statistic for testing* $H_0 : \alpha = 0$ *or* $\mu = \mu_0$. *Suppose that Conditions* $A1 - A7$ *hold, then under the null distribution* $f(\theta; \mu_0, \kappa_0)$, *the limiting distribution of* M_n *is* χ_1^2.

COROLLARY 1 *Let* $\theta_1, \ldots, \theta_n$ *be a random sample from the mixture population* $(1-\alpha)f(\theta; \mu_0, \kappa) + \alpha f(\theta; \mu, \kappa)$, *where* $0 \leq \alpha \leq 1$, $\mu \in \mathbf{T}$, μ_0 *is on the boundary of* $\mathbf{T}$ *and* $\kappa \in \mathbf{B}$. *Let* M_n *be (twice) the MLRT statistic for testing* $H_0 : \alpha = 0$ *or* $\mu = \mu_0$. *Suppose that Conditions* $A1 - A7$ *hold. Then under the null distribution* $f(\theta; \mu_0, \kappa_0)$, *the limiting distribution of* M_n *is* $1/2\chi_0^2 + 1/2\chi_1^2$.

5. Simulation Study and Real Data Examples

The purpose of the simulation study is to examine the proposed asymptotic null distribution of the MLRT statistic. Samples of size n(=50, 100, 200, 500) are generated from a single von Mises distribution with mean direction zero and concentration parameter κ(=1, 2, 3, 4). For each set of sample size n and concentration parameter κ, the empirical null distribution of M_n is obtained using 10,000 replications. Three nominal significance levels 10%, 5% and 1% are examined. We used "optim" function in R to maximize the modified log-likelihood function. Several different initial values were tried to increase the chance of locating the global maximum.

The simulated null rejection rates of the MLRT with the level of modification $C = 1$ are presented in Table 1. We find that when $C = 1$ with moderate sample sizes the simulated null rejection rates are quite close to the values given by the asymptotic theory. Figure 1 gives Q-Q plots for $\kappa = 3$ with $C = 1$. The agreement between the simulated quantiles and those of χ_1^2 becomes satisfactory when the sample size is around 200.

We now apply the MLRT to two real data examples. The first one is from a genetic background. Storch et al. (2002) studied the circadian gene expression in mice liver and heart. In the study, mice were synchronized to a 12-hour light/dark cycle for more than two weeks, then placed in a constant dim light for more than 42 hours. The tissue samples were collected from sacrificed mice at 4-hour intervals over two circadian cycles. It was found that "the liver and heart circadian gene sets revealed very little overlap, with only 52 genes in common". Liu et al. (2006) estimated the phase angles of 48 cycling transcripts in the two tissues using a random-period model. Four transcripts were excluded from the analysis due to lack of fit of the model. We analyze the angular difference data of the 48 cycling transcripts (heart minus liver) under (). The MLRT is used to test whether there exists a subset of the 48 genes having unequal phase angles in the two tissues. The modified MLEs are $\hat{\alpha}^* = 0.30$, $\hat{\mu}^* = 1.99$, $\hat{\kappa}^* = 1.74$, and $\hat{\kappa}_0^* = 0.83$. The MLRT statistic is found to be 2.18 with $C = 1$. According the χ_1^2 limiting distribution, the asymptotic p-value is 0.14 which suggests lack of evidence to reject $H_0 : M(0, \kappa)$. In other words, the two-component von Mises mixture distribution with common concentration parameter does not provide a statistically significantly better fit to the data. See Figure 2 for the comparison of a kernel density estimate and the density of the estimated two-component von Mises mixture distribution with bandwidth 15. Mixture models with more components may have good appeals, but the small data set lacks sufficient information to be conclusive. Liu et al. (2006) fitted the data with a two-component von Mises mixture model with unequal κ. The test problem with unequal κ is mathematically more complex and still unsolved which is worth further study.

The second data set has geological backgound. Information about the flow directions of ancient rivers (paleoflow direction) helps scientists better understand how certain rock units are oriented, which in turn leads to more efficient exploration of natural resources and better understanding of landscape development and climate change. Primary bedforms are usually used to interpret the paleoflow direction, since the orientation of the foreset lamination of the bedforms parallels to the current direction. However, bedforms are often masked or destroyed by various physical and chemical processes, scientists then have to analyze other available data to obtain information on the paleoflow direction. Morris et al. (1996) proposed the use of the orientation of elongate bones as additional information to identify the paleoflow direction. Consequently, it is of real importance to test the hypothesis on whether the orientation of elongate bones is consistent with paleoflow direction. Dinosaur National Monument and Dry Mesa Dinosaur Quarry are two ideal quarries to be used for comparison of directions of elongate bone and paleoflow, since both dinosaur bone and well-preserved bedforms exist. As pointed out in Grimshaw et al. (2001), elongate bones can be classified into two categories: symmetrical and asymmetrical. Symmetrical bones tend to orient themselves vertical to the paleoflow direction, while asymmetrical bones, which display additional bone mass on only one end, tend to orient themselves parallel to the paleoflow direction. Consequently, Grimshaw et al. (2001) proposed the use of mixture of von Mises distributions to model the bone directional data for the purpose of statistical hypothesis test. Their analysis suggested that one of the mean directions in the von Mises mixture distribution is consistent with the paleoflow direction. The result hence supports the use of dinosaur bone orientations to estimate paleoflow direction when the bedforms are not visible.

In this paper, we use the data to test the hypothesis whether the second category of the bone in fact exists. It is seen that the direction of the asymmetrical bones can be treated as known, because the estimated paleflow direction is available. For this purpose, let us first introduce more details about the data. The measurements on elongate dinosaur bones are axial data with period π, since there is no reason to make a distinction of two ends of the fossil bone. In order to use the vectorial probability models, one can double the angles modulo 2π. The values of the transformed axial data then range from 0 to 2π. Dinosaur National Monument and Dry Mesa Dinosaur Quarry have 444 and 555 dinosaur bones direction measurements, respectively. The paleoflow directions for two locations, which are estimated from the well-preserved primary bedforms, are 1.046π and 0.820π in transformed axial units.

We applied the MLRT to both sets of dinosaur bone data. For Dinosaur National Monument data with $\mu_0 = 1.046\pi$, the modified MLEs are $\hat{\alpha}^* = 0.29$, $\hat{\mu}^* = 1.93\pi$, $\hat{\kappa}^* = 1.45$. The MLRT statistic is found to be 26.77 with $C = 1$, which suggests strong evidence to reject the unicomponent von Mises distribution. For Dry Mesa Dinosaur Quarry data with $\mu_0 = 0.820\pi$, the modified MLEs are $\hat{\alpha}^* = 1.00$, $\hat{\mu}^* = 0.75\pi$, $\hat{\kappa}^* = 0.20$. The MLRT statistic is found to be 0.55 with $C = 1$. According the χ^2_1 limiting distribution, the asymptotic p-value is 0.46, which suggests lack of evidence to reject

$H_0 : M(\mu_0, \kappa)$. Figures 1 and 2 in Grimshaw et al. (2001) present the nonparametric density estimate of the two data sets. For Dinosaur National Monument data, the nonparametric density curve is clearly bimodal, however, for Dry Mesa Dinosaur Quarry data, there is no apparent indication of mixture model.

6. Summary Comments

In this paper, we investigate the use of the MLRT for homogeneity in a mixture of directional distributions. In particular, we consider the test for a unicomponent von Mises distribution against a two-component von Mises mixture with common unknown concentration parameter. We find that the MLRT has a simple χ_1^2 null limiting distribution and is very easy to use in applications. This is the very first result on the use of the modified likelihood approach in finite mixture models for circular data analysis. We expect the MLRT shares many other nice properties when applied to directional data.

We also extend the result to a mixture model with a general parametric kernels. There are a lot of examples in the literature which are special cases of this general formulation. For instance, the one parameter mixture of exponential distribution considered in Slud (1997), the one parameter Gamma mixture considered in Liu et al. (2003) and the one parameter mixture of location shift kernel in Devline et al. (2000).

For the von Mises kernel, a more general two-component mixture is

$$(1 - \alpha)M(\mu_1, \kappa_1) + \alpha M(\mu_2, \kappa_2),$$

where $0 \leq \alpha \leq 1$, $|\mu_1| \leq \pi$, $|\mu_2| \leq \pi$, $\kappa_1 \geq 0$ and $\kappa_2 \geq 0$. Interestingly, the likelihood function of such mixture model is also unbounded similar to the Normal mixture in linear data. Therefore, the likelihood method can not be directly applied. The modified likelihood approach provides an attractive alternative. A theory is yet to be developed and the problem is still under investigation.

Acknowledgment

The research was partially supported by grants from the Natural Sciences and Engineering Research Council of Canada.

References

Abramowitz, M. and Stegun, I. A. (1965). Handbook of mathematical functions. Dover, New York.

Adler, R. J. (1990). An introduction to continuity, extrema, and related topics for general Gaussian processes. *IMS Lecture Notes - Monograph Series*, **12**. Institute of Mathematical Statistics, Hayward.

Billingsley, P. (1968). Convergence of probability measures. Wiley, New York.

Chen, H. and Chen, J. (2001). The likelihood ratio test for homogeneity in the finite mixture models. *Canadian Journal of Statistics*, **29**, 201-215.

Chen, H. and Chen, J. (2003). Tests for homogeneity in normal mixtures with presence of a structural parameter. *Statistica Sinica*, **13**, 351-365.

Chen, H., Chen, J. and Kalbfleisch, J. D. (2000). A modified likelihood ratio test for homogeneity in the finite mixture models. Working Paper 2000-01 Department of Statistics and Actuarial Science, University of Waterloo.

Chen, H., Chen, J. and Kalbfleisch, J. D. (2001). A modified likelihood ratio test for homogeneity in finite mixture models. *Journal of the Royal Statistical Society, Series B*, **63**, no. 1, 19-29.

Chen, H., Chen, J. and Kalbfleisch, J. D. (2004). Testing for a finite mixture model with two components. *Journal of the Royal Statistical Society, Series B*, **66**, no. 1, 95-115.

Chen, J. (1998). Penalized likelihood ratio test for finite mixture models with multinomial observations. *Canadian Journal of Statistics*, **26**, 583-599.

Chen, J. and Kalbfleisch, J. D. (2005). Modified likelihood ratio test in finite mixture models with a structural parameter. *Journal of Statistical Planning and Inference*, **129**, 93-107.

Dacunha-Castelle, D. and Gassiat, E. (1999). Testing the order of a model using locally conic parametrization: population mixtures and stationary ARMA processes. *Annals of Statistics*, **27**, 1178-1209.

Devlin, B., Roeder, K. and Wasserman, L. (2000). Genomic control for association studies: a semiparametric test to detect excess-haplotype sharing. *Biostatistics*, 369-387.

Fraser, M. D., Hsu, Y. S. and Walker, J. J. (1981). Identifiability of finite mixtures of von Mises distributions. *The Annals of Statistics*, **9**, 1130-1131.

Grimshaw, S. D., Whiting, D. G. and Morris, T. H. (2001). Likelihood ratio tests for a mixture of two von Mises distributions. *Biometrics*, **57**, 260-265.

Holzmann, H., Munk, A. and Stratmann, B. (2004). Identifiability of finite mixtures - with applications to circular distributions. *Sankhya*, **66**, 440-450.

Phase analysis of circadian-related genes in two tissues Delong Liu, Shyamal D Peddada, Leping Li, Clarice R Weinberg BMC Bioinformatics 2006, 7:87 (23 February 2006)

Liu, D., Peddada, S. D., Li, L., Weinberg, C. R. (2006). Phase analysis of circadian-related genes in two tissues. *BMC Bioinformatics*, **7**:87.

Liu, X., Pasarica, C. and Shao, Y. Z. (2003). Testing homogeneity in gamma mixture models. *Scandinavian Journal of Statistics*, 30, 227-239.

Liu, X. and Shao, Y. Z. (2003). Asymptotics for likelihood ratio tests under loss of identifiability. *Ann. Statist.*, **31**, 807-832.

Mardia, K. V. and Jupp, P. E. (2000). Directional statistics. John Wiley and Sons.

Morris, T. H., Richmond, D. R., and Grimshaw, S. D. (1996). Orientation of dinosaur bones in riverine environments: insights into sedimentary dynamics and taphonomy. *The Continental Jurassic*, Morales, M., ed., 521-530. Museum of Northern Arizona, Flagstaff, Arizona.

Panda, S., Antoch, M. P., Miller, B. H., Su, A. I., Schook, A. B., Straume, M., Schultz, P. G., Kay, S. A., Takahashi, J. S., Hogenesch, J. B. (2002). Coordinated transcription of key pathways in the mouse by the circadian clock. *Cell*, **109**, 307-320.

Rubin H. (1956). Uniform convergence of random functions with applications to statistics. *Annals of Mathematical Statistics*, **27**, 200-203.

Slud, E. (1997). Testing for imperfect debugging in software reliability. *Scandinavian Journal of Statistics*, **24**, 555-572.

Storch, K. F., Lipan, O., Leykin, I., Viswanathan, N., Davis, F. C., Wong, W. H., Weitz, C. J. (2002). Extensive and divergent circadian gene expression in liver and heart. *Nature*, **417**, 78-83.

von Mises, R. (1918). *Über* die "ganzzahligkeit" der atomgewichte und verwandte fragen. Phys. Z., **19**, 490-500.

Wald, A. (1949). Note on the consistency of the maximum likelihood estimate. *Annals of Mathematical Statistics*, **20**, 595-601.

Zhu, H. T. and Zhang, H. P. (2004). Hypothesis testing in mixture regression models. *Journal of the Royal Statistical Society, Series B*, **66**, no. 1, 3-16.

Appendix

We first set up two propositions needed in the proofs of the theorems. We employ here the so-called "sandwich method" to derive the limiting distribution of the LRT statistic R_n under the null hypothesis. The basic idea of the sandwich method is to establish an upper bound to R_n first and then justify that the upper bound can be attained at some choice of parameter values. In light of Lemma 2, we can restrict κ within an interval $[\kappa_0 - \delta, \kappa_0 + \delta]$ for some small constant $0 < \delta < 1$ for theoretical derivations.

Let $r_n(\alpha, \mu, \kappa) = 2\{l_n(\alpha, \mu, \kappa) - l_n(0, 0, \hat{\kappa}_0)\}$ be a likelihood function. It is seen that $R_n = r_n(\hat{\alpha}, \hat{\mu}, \hat{\kappa})$. Write $r_n(\alpha, \mu, \kappa) = r_{1n}(\alpha, \mu, \kappa) + r_{2n}$, where $r_{1n}(\alpha, \mu, \kappa) = 2\{l_n(\alpha, \mu, \kappa) - l_n(0, 0, \kappa_0)\}$ and $r_{2n} = 2\{l_n(0, 0, \kappa_0) - l_n(0, 0, \hat{\kappa}_0)\}$.

We first study $r_{1n}(\alpha, \mu, \kappa)$. Express $r_{1n}(\alpha, \mu, \kappa) = 2\sum_{i=1}^n \log(1 + \delta_i)$, where

$$\delta_i = (1-\alpha)\Big\{\frac{f(\theta_i; 0, \kappa)}{f(\theta_i; 0, \kappa_0)} - 1\Big\} + \alpha\Big\{\frac{f(\theta_i; \mu, \kappa)}{f(\theta_i; 0, \kappa_0)} - 1\Big\}. \tag{8}$$

Define

$$\begin{aligned}
U_i(\kappa) &= \frac{1}{\kappa - \kappa_0}\Big\{\frac{f(\theta_i; 0, \kappa)}{f(\theta_i; 0, \kappa_0)} - 1\Big\} = \frac{1}{\kappa - \kappa_0}\Big[\frac{I_0(\kappa_0)}{I_0(\kappa)}\exp\{(\kappa - \kappa_0)\cos\theta_i\} - 1\Big]; \\
Y_i(\mu, \kappa) &= \frac{1}{\mu}\Big\{\frac{f(\theta_i; \mu, \kappa)}{f(\theta_i; 0, \kappa_0)} - \frac{f(\theta_i; 0, \kappa)}{f(\theta_i; 0, \kappa_0)}\Big\} \\
&= \frac{I_0(\kappa_0)}{\mu I_0(\kappa)}[\exp\{\kappa\cos(\theta_i - \mu) - \kappa_0\cos\theta_i\} - \exp\{(\kappa - \kappa_0)\cos\theta_i\}],
\end{aligned}$$

and $Y_i(0, \kappa)$ and $U_i(\kappa_0)$ be their continuity limits. For convenience of notation, we put $Y_i(\mu) = Y_i(\mu, \kappa_0)$ and $U_i = U_i(\kappa_0)$. With these definitions, δ_i can be expressed as

$$\delta_i = (\kappa - \kappa_0)U_i + \alpha\mu Y_i(\mu) + \epsilon_{in}, \tag{9}$$

where $\epsilon_{in} = (\kappa - \kappa_0)\{U_i(\kappa) - U_i\} + \alpha\mu\{Y_i(\mu, \kappa) - Y_i(\mu)\}$. The following proposition assesses the stochastic order of the relevant processes formed by $U_i(\kappa)$ and $Y_i(\mu, \kappa)$.

Proposition 1 *Under the null distribution* $M(0,\kappa_0)$, *for* $\kappa \in [\kappa_0-\delta, \kappa_0+\delta]$ *and* $|\mu| \leq \pi$, *the following processes are tight*

$$\begin{aligned} U_n^*(\kappa) &= n^{-1/2} \sum \{U_i(\kappa) - U_i\}/(\kappa - \kappa_0), \\ Y_n^*(\mu) &= n^{-1/2} \sum \{Y_i(\mu) - Y_i(0)\}/\mu, \\ Y_n^*(\mu,\kappa) &= n^{-1/2} \sum \{Y_i(\mu,\kappa) - Y_i(\mu)\}/(\kappa - \kappa_0). \end{aligned}$$

Proof. In light of Billingsley (1968, p.95), we need to verify the Lipschitz conditions

$$E\{U_n^*(\kappa_1) - U_n^*(\kappa_2)\}^2 \leq B(\kappa_1 - \kappa_2)^2$$

$$E\{Y_n^*(\mu_1) - Y_n^*(\mu_2)\}^2 \leq B(\mu_1 - \mu_2)^2$$

$$E\{Y_n^*(\mu_1,\kappa_1) - Y_n^*(\mu_2,\kappa_2)\}^2 \leq B[(\mu_1 - \mu_2)^2 + (\kappa_1 - \kappa_2)^2],$$

for some constant B. Consider the following functions

$$\frac{U_i(\kappa) - U_i}{\kappa - \kappa_0}, \quad \frac{Y_i(\mu) - Y_i(0)}{\mu} \quad \text{and} \quad \frac{Y_i(\mu,\kappa) - Y_i(\mu)}{\kappa - \kappa_0}.$$

The Lipschitz condition is satisfied if the derivatives of the above functions have bounded second moments uniformly in μ and κ. This is obvious since their second moments are continuous in μ and κ inside a compact parameter space. □

By the inequality $2\log(1+x) \leq 2x - x^2 + (2/3)x^3$, we have

$$r_{1n}(\alpha,\mu,\kappa) = 2\sum_{i=1}^{n} \log(1+\delta_i) \leq 2\sum_{i=1}^{n} \delta_i - \sum_{i=1}^{n} \delta_i^2 + \frac{2}{3}\sum_{i=1}^{n} \delta_i^3.$$

Proposition 2 *Under the null model* $M(0,\kappa_0)$, *we have*

$$\begin{aligned} \sum_{i=1}^{n} \delta_i - \sum_{i=1}^{n} \{(\kappa-\kappa_0)U_i + \alpha\mu Y_i(\mu)\} &= n^{1/2}\{(\kappa-\kappa_0)^2 + (\alpha\mu)^2\}O_p(1), && (10) \\ \sum_{i=1}^{n} \delta_i^2 - \sum_{i=1}^{n} \{(\kappa-\kappa_0)U_i + \alpha\mu Y_i(\mu)\}^2 &= n\{(\kappa-\kappa_0)^2 + (\alpha\mu)^2\}|\kappa-\kappa_0|O_p(1), && (11) \\ \sum_{i=1}^{n} \delta_i^3 - \sum_{i=1}^{n} \{(\kappa-\kappa_0)U_i + \alpha\mu Y_i(\mu)\}^3 &= n\{(\kappa-\kappa_0)^2 + (\alpha\mu)^2\}^2 O_p(1). && (12) \end{aligned}$$

Proof. By Proposition 1, we have

$$\sup_{\kappa\in[\kappa_0-\delta,\kappa_0+\delta]} U_n^*(\kappa) = O_p(1), \;\; \sup_{|\mu|\le\pi} Y_n^*(\mu) = O_p(1), \text{ and } \sup_{\kappa\in[\kappa_0-\delta,\kappa_0+\delta],\ |\mu|\le\pi} Y_n^*(\mu,\kappa) = O_p(1).$$

Hence,

$$\begin{aligned}\sum_{i=1}^n \epsilon_{in} &= (\kappa-\kappa_0)\sum_{i=1}^n\{U_i(\kappa)-U_i\} + \alpha\mu\sum_{i=1}^n\{Y_i(\mu,\kappa)-Y_i(\mu)\}\\ &= n^{1/2}\{(\kappa-\kappa_0)^2+(\alpha\mu)^2\}O_p(1).\end{aligned}$$

This proves (10) in the proposition. For the square term, we have

$$\sum_{i=1}^n \delta_i^2 - \sum_{i=1}^n\{(\kappa-\kappa_0)U_i+\alpha\mu Y_i(\mu)\}^2 = \sum_{i=1}^n \epsilon_{in}^2 + 2\sum_{i=1}^n \epsilon_{in}\{(\kappa-\kappa_0)U_i+\alpha\mu Y_i(\mu)\}.$$

Note that

$$\sum_{i=1}^n \epsilon_{in}^2 \le n(\kappa-\kappa_0)^2\left[(\kappa-\kappa_0)^2+(\alpha\mu)^2\right]O_p(1) \quad \text{and} \tag{13}$$

$$\sum_{i=1}^n \epsilon_{in}\{(\kappa-\kappa_0)U_i+\alpha\mu Y_i(\mu)\} \le n\{(\kappa-\kappa_0)^2+(\alpha\mu)^2\}|\kappa-\kappa_0|O_p(1). \tag{14}$$

Combining (13) and (14), conclusion (11) therefore follows.

Similarly, for the cubic term of δ_i we have

$$\sum_{i=1}^n \delta_i^3 - \sum_{i=1}^n\{(\kappa-\kappa_0)U_i+\alpha\mu Y_i(\mu)\}^3 = n\{(\kappa-\kappa_0)^2+\alpha^2\mu^2\}^2 O_p(1).$$

Conclusion (12) follows. □

Proof of Theorem 1. By Proposition 2, we have

$$\begin{aligned}r_{1n}(\alpha,\mu,\kappa) \le\; & 2\sum_{i=1}^n\{(\kappa-\kappa_0)U_i+\alpha\mu Y_i(\mu)\} - \sum_{i=1}^n\{(\kappa-\kappa_0)U_i+\alpha\mu Y_i(\mu)\}^2\\ &+\frac{2}{3}\sum_{i=1}^n\{(\kappa-\kappa_0)U_i+\alpha\mu Y_i(\mu)\}^3 + n^{1/2}\{(\kappa-\kappa_0)^2+\alpha^2\mu^2\}O_p(1)\\ &+n\{(\kappa-\kappa_0)^2+\alpha^2\mu^2\}|\kappa-\kappa_0|O_p(1)+n\{(\kappa-\kappa_0)^2+\alpha^2\mu^2\}^2O_p(1).\end{aligned} \tag{15}$$

Note that, under the null distribution $M(0, \kappa_0)$,

$$E\{U_i Y_i(\mu)\} = A(\kappa_0)(\cos\mu - 1)/\mu \quad \text{and} \quad E(U_i^2) = 1 - A(\kappa_0)/\kappa_0 - A^2(\kappa_0), \tag{16}$$

where $A(\kappa_0)$ is defined by $I_1(\kappa_0)/I_0(\kappa_0)$ in (2). Let

$$Z_i(\mu) \;\; = \;\; Y_i(\mu) - \frac{E\{U_i Y_i(\mu)\}}{E(U_i^2)} U_i = Y_i(\mu) - \frac{A(\kappa_0)(\cos\mu - 1)}{\mu\{1 - A(\kappa_0)/\kappa_0 - A^2(\kappa_0)\}} U_i.$$

Then $(\kappa - \kappa_0)U_i + \alpha\mu Y_i(\mu) = t_1 U_i + t_2 Z_i(\mu)$, where $t_2 = \alpha\mu$ and

$$t_1 = \kappa - \kappa_0 + \frac{A(\kappa_0)(\cos\mu - 1)}{\mu\{1 - A(\kappa_0)/\kappa_0 - A^2(\kappa_0)\}} t_2. \tag{17}$$

It is easy to verify that $E\{U_i Z_i(\mu)\} = 0$. That is U_i and $Z_i(\mu)$ are orthogonal for all μ. Note that $n^{-1}\sum_{i=1}^n \{t_1 U_i + t_2 Z_i(\mu)\}^2$ converges uniformly to a positive definite quadratic form in t_1 and t_2 and $n^{-1}\sum_{i=1}^n \{|U_i|^3 + |Z_i(\mu)|\}^3 = O_p(1)$ uniformly in $|\mu| \le \pi$. Thus

$$\frac{\sum_{i=1}^n |t_1 U_i + t_2 Z_i(\mu)|^3}{\sum_{i=1}^n \{t_1 U_i + t_2 Z_i(\mu)\}^2} \le (|t_1| + |t_2|) O_p(1)$$

and

$$\sum_{i=1}^n \{t_1 U_i + t_2 Z_i(\mu)\}^2 \ge \lambda n (t_1^2 + t_2^2)$$

in probability for some $\lambda > 0$ uniformly in μ. Observe that uniformly in μ,

$$\frac{A(\kappa_0)(\cos\mu - 1)}{\mu\{1 - A(\kappa_0)/\kappa_0 - A^2(\kappa_0)\}} = O(1).$$

By (17), we have $(\kappa - \kappa_0)^2 = \{t_1 - t_2 O(1)\}^2 \le (t_1^2 + t_2^2) O(1)$. Thus

$$\begin{aligned} n^{1/2}\{(\kappa - \kappa_0)^2 + \alpha^2\mu^2\} O_p(1) &\le n^{1/2}(t_1^2 + t_2^2) O_p(1) = o_p\{\sum_{i=1}^n [t_1 U_i + t_2 Z_i(\mu)]^2\}, \\ n\{(\kappa - \kappa_0)^2 + \alpha^2\mu^2\} |\kappa - \kappa_0| O_p(1) &\le (|t_1| + |t_2|) O_p\{\sum_{i=1}^n [t_1 U_i + t_2 Z_i(\mu)]^2\} \quad \text{and} \\ n\{(\kappa - \kappa_0)^2 + \alpha^2\mu^2\}^2 O_p(1) &\le (t_1^2 + t_2^2) O_p\{\sum_{i=1}^n [t_1 U_i + t_2 Z_i(\mu)]^2\}. \end{aligned}$$

It follows that (15) can be rewritten as

$$\begin{aligned} r_{1n}(\alpha,\mu,\kappa) &\le 2\sum_{i=1}^{n}\{t_1U_i+t_2Z_i(\mu)\} \\ &\quad -\sum_{i=1}^{n}\{t_1U_i+t_2Z_i(\mu)\}^2\{1+(|t_1|+|t_2|+t_1^2+t_2^2)O_p(1)+o_p(1)\}. \end{aligned}$$

Since U_i and $Z_i(\mu)$ are orthogonal, the above inequality can be further reduced to

$$\begin{aligned} r_{1n}(\alpha,\mu,\kappa) &\le 2\sum_{i=1}^{n}\{t_1U_i+t_2Z_i(\mu)\} \\ &\quad -\sum_{i=1}^{n}\{t_1^2U_i^2+t_2^2Z_i^2(\mu)\}\{1+(|t_1|+|t_2|+t_1^2+t_2^2)O_p(1)+o_p(1)\}. \end{aligned} \tag{18}$$

Let us now restrict our attention to a small neighborhood of $(t_1,t_2)=(0,0)$ as suggested by the consistency results of the MLEs in Lemma 2. Consequently, we may regard t_1 and t_2 as $o_p(1)$. Inequality (18) then becomes

$$r_{1n}(\alpha,\mu,\kappa) \le 2\sum_{i=1}^{n}\{t_1U_i+t_2Z_i(\mu)\}-\sum_{i=1}^{n}\{t_1^2U_i^2+t_2^2Z_i^2(\mu)\}\{1+o_p(1)\}. \tag{19}$$

Furthermore, the right-hand side of (19) is asymptotically less than or equal to the maximum of the following quadratic function

$$Q(t_1,t_2)=2\sum_{i=1}^{n}\{t_1U_i+t_2Z_i(\mu)\}-\sum_{i=1}^{n}\{t_1^2U_i^2+t_2^2Z_i^2(\mu)\}.$$

Note that for any fixed μ, t_2 has the same sign as μ and $Q(t_1,t_2)$ is maximized at $t_1=\tilde{t}_1$ and $t_2=\tilde{t}_2$ with

$$\tilde{t}_1=\frac{\sum U_i}{\sum U_i^2},\quad \tilde{t}_2=\frac{[\mathrm{sgn}(\mu)\sum Z_i(\mu)]^+}{\sum Z_i^2(\mu)}, \tag{20}$$

where $\mathrm{sgn}(\mu)$ is the sign function. Thus

$$r_{1n}(\hat{\alpha},\hat{\mu},\hat{\kappa}) \le \frac{\{\sum U_i\}^2}{\sum U_i^2}+\sup_{|\mu|\le\pi}\frac{\{[\mathrm{sgn}(\mu)\sum Z_i(\mu)]^+\}^2}{\sum Z_i^2(\mu)}+o_p(1). \tag{21}$$

We have established the asymptotic upper bound for r_{1n}. Next, we prove that the asymptotic upper bound can be attained at a set of parameter values. Let $\epsilon>0$ be any fixed small number. For any fixed $\epsilon \leq |\mu| \leq \pi$, let $\tilde{\kappa}(\mu)$ and $\tilde{\alpha}(\mu)$ be the values determined by (20). Consider the Taylor series expansion

$$r_{1n}(\tilde{\alpha}(\mu), \mu, \tilde{\kappa}(\mu)) = 2\sum_{i=1}^{n} \tilde{\delta}_i - \sum_{i=1}^{n} \tilde{\delta}_i^{\,2}(1+\tilde{\eta}_i)^{-2},$$

where $|\tilde{\eta}_i| < |\tilde{\delta}_i|$ and $\tilde{\delta}_i$ is equal to δ_i in (8) with $\kappa = \tilde{\kappa}(\mu)$ and $\alpha = \tilde{\alpha}(\mu)$. Attributing to bounding away from 0, the solution $\tilde{\alpha}(\mu)$ is feasible, so that $\tilde{\alpha}(\mu) = O_p(n^{-1/2})$ and $\tilde{\kappa}(\mu) - \kappa_0 = O_p(n^{-1/2})$ uniformly in $\epsilon \leq |\mu| \leq \pi$. Since $U_i(\mu)$ and $Y_i(\mu, \kappa)$ are bounded functions for $|\mu| \leq \pi$, $|\theta_i| \leq \pi$ and $\kappa \in [\kappa_0 - \delta, \kappa_0 + \delta]$, we have $\max_{1\leq i\leq n} |\tilde{\delta}_i| = O_p(n^{-1/2}) = o_p(1)$. It follows that uniformly in μ, $\max |\tilde{\eta}_i| = o_p(1)$. Then we can get

$$\begin{aligned} r_{1n}(\tilde{\alpha}(\mu), \mu, \tilde{\kappa}(\mu)) &= 2\sum_{i=1}^{n} \tilde{\delta}_i - \sum_{i=1}^{n} \tilde{\delta}_i^{\,2}\{1 + o_p(1)\} \\ &= \frac{\{\sum U_i\}^2}{\sum U_i^2} + \sup_{\epsilon\leq|\mu|\leq\pi} \frac{\{[\mathrm{sgn}(\mu)\sum Z_i(\mu)]^+\}^2}{\sum Z_i^2(\mu)} + o_p(1). \end{aligned}$$

Thus, for any fixed $\epsilon > 0$,

$$r_{1n}(\hat{\alpha}, \hat{\mu}, \hat{\kappa}) \geq \frac{\{\sum U_i\}^2}{\sum U_i^2} + \sup_{\epsilon\leq|\mu|\leq\pi} \frac{\{[\mathrm{sgn}(\mu)\sum Z_i(\mu)]^+\}^2}{\sum Z_i^2(\mu)} + o_p(1). \tag{22}$$

Note that r_{2n} has an ordinary quadratic approximation, i.e.,

$$r_{2n} = -\frac{\{\sum U_i\}^2}{\sum U_i^2} + o_p(1). \tag{23}$$

Therefore for any fixed $\epsilon > 0$,

$$\sup_{\epsilon\leq|\mu|\leq\pi} \frac{\{[\mathrm{sgn}(\mu)\sum Z_i(\mu)]^+\}^2}{\sum Z_i^2(\mu)} + o_p(1) \leq R_n \leq \sup_{|\mu|\leq\pi} \frac{\{[\mathrm{sgn}(\mu)\sum Z_i(\mu)]^+\}^2}{\sum Z_i^2(\mu)} + o_p(1). \tag{24}$$

By the uniform strong law of large numbers, $n^{-1}\sum_{i=1}^{n} Z_i^2(\mu) \to EZ_1^2(\mu)$ almost surely and uniformly in $|\mu| \leq \pi$. Thus we can rewrite (24) as

$$\sup_{\epsilon\leq|\mu|\leq\pi} \frac{\{[\mathrm{sgn}(\mu)\sum Z_i(\mu)]^+\}^2}{nEZ_1^2(\mu)} + o_p(1) \leq R_n \leq \sup_{|\mu|\leq\pi} \frac{\{[\mathrm{sgn}(\mu)\sum Z_i(\mu)]^+\}^2}{nEZ_1^2(\mu)} + o_p(1). \tag{25}$$

Notice that by the tightness of the process $Y_n^*(\mu)$ the process

$$\{nEZ_1^2(\mu)\}^{-\frac{1}{2}} \sum_{i=1}^{n} Z_i(\mu), \quad |\mu| \leq \pi$$

converges weakly to a Gaussian process $\xi(\mu)$. Direct calculation of the mean and covariance of $Z_i(\mu)$ yields that the Gaussian process $\xi(\mu)$ has mean 0, standard deviation 1 and the autocorrelation function

$$\rho(s,t) = \mathrm{sgn}(st) \frac{g(s,t)}{\{g(s,s)g(t,t)\}^{\frac{1}{2}}}, \qquad \text{for } s,t \neq 0,$$

where $g(s,t) = E\{Z_1(s)Z_1(t)\}$.

By letting $n \to \infty$ and then $\epsilon \to 0$ in (25), we find R_n converges in probability to $\sup_{|\mu|\leq\pi} \zeta^+(\mu)$, where the Gaussian process $\zeta(\mu) = \mathrm{sgn}(\mu)\xi(\mu)$ has the mean 0 and autocorrelation function $\rho(s,t)$ given in (5). □

Remark. The calculation of the autocorrelation function $g(s,t)$ is as follows.

$$E\{Z_1(s)Z_1(t)\} = E\{Y_1(s)Y_1(t)\} - \frac{E\{U_1Y_1(s)\}E\{U_1Y_1(t)\}}{E(U_1^2)},$$

where by (16)

$$\begin{aligned} \frac{E\{U_1Y_1(s)\}E\{U_1Y_1(t)\}}{E(U_1^2)} &= \frac{A^2(\kappa_0)(\cos s - 1)(\cos t - 1)}{st\{1 - A(\kappa_0)/\kappa_0 - A^2(\kappa_0)\}} \quad \text{and} \\ E\{Y_1(s)Y_1(t)\} &= \frac{1}{st} E[\exp\{\kappa_0 \cos(\theta_1 - s) + \kappa_0 \cos(\theta_1 - t) - 2\kappa_0 \cos\theta_1\}] - \frac{1}{st}. \end{aligned}$$

By the trigonometric identity $\cos(x-y) = \cos x \cos y + \sin x \sin y$, we have

$$\cos(\theta_1 - s) + \cos(\theta_1 - t) - \cos\theta_1 = \cos(\theta_1 - \eta)\{(\cos s + \cos t - 1)^2 + (\sin s + \sin t)^2\}^{\frac{1}{2}} \quad (26)$$

where $\cos\eta = (\cos s + \cos t - 1)\{(\cos s + \cos t - 1)^2 + (\sin s + \sin t)^2\}^{-\frac{1}{2}}$. Using the identity in (26), we obtain

$$E\{Y_1(s)Y_1(t)\} = \frac{1}{st} \frac{I_0[\kappa_0\{(\cos s + \cos t - 1)^2 + (\sin s + \sin t)^2\}^{\frac{1}{2}}]}{I_0(\kappa_0)} - \frac{1}{st}.$$

Hence, the covariance of $Z_1(s)$ and $Z_1(t)$ is

$$g(s,t) = \frac{1}{st}\Big[\frac{I_0[\kappa_0\{(\cos s+\cos t-1)^2+(\sin s+\sin t)^2\}^{\frac{1}{2}}]}{I_0(\kappa_0)} - 1 - \frac{A^2(\kappa_0)(\cos s-1)(\cos t-1)}{1-A(\kappa_0)/\kappa_0 - A^2(\kappa_0)}\Big].$$

Proof of Theorem 2. Since $pl_n(1,0,\kappa) = l_n(0,0,\kappa)$, $\hat\kappa_0^*$ is in fact equal to $\hat\kappa_0$, we have $M_n = r_n(\hat\alpha^*,\hat\mu^*,\hat\kappa^*) + 2C\log(\hat\alpha^*)$. As a first step, we show that $\log(\hat\alpha^*) = O_p(1)$. Since $0 \le M_n \le R_n = O_p(1)$, we conclude that $M_n = O_p(1)$. In addition, $0 \le M_n - C\log(\hat\alpha^*) \le R_n$, so $M_n - C\log(\hat\alpha^*) = O_p(1)$ which implies $\log(\hat\alpha^*) = O_p(1)$. Hence we can consider the problem by restricting α within the closed interval $[\delta_0, 1]$, where $0 < \delta_0 < 1$ and prove the consistency of the modified MLEs, i.e., $\hat\mu^* \xrightarrow{p} 0$ and $\hat\kappa^* \xrightarrow{p} \kappa_0$. For a similar proof, see Chen et al. (2000).

Recall $Y_i(\mu) = Y_i(\mu,\kappa_0)$ and let $Y_i = Y_i(0)$. It is easy to see that U_i and $Y_i = \kappa_0 \sin\theta_i$ are orthogonal. Rewrite δ_i in (9) as

$$\delta_i = (\kappa-\kappa_0)U_i + \alpha\mu Y_i + \epsilon_{in}^*. \tag{27}$$

The remainder is $\epsilon_{in}^* = (\kappa-\kappa_0)\{U_i(\kappa) - U_i(\kappa_0)\} + \alpha\mu\{Y_i(\mu,\kappa) - Y_i(\mu) + Y_i(\mu) - Y_i\}$. In light of Proposition 1, for $\alpha \in [\delta_0, 1]$,

$$\sum \epsilon_{in}^* = n^{1/2}(\kappa-\kappa_0)^2 O_p(1) + n^{1/2}\alpha\mu(\kappa-\kappa_0)O_p(1) + n^{1/2}\alpha\mu^2 O_p(1).$$

Using the similar arguments as in the proof of Theorem 1, we get

$$r_{1n}(\hat\alpha^*,\hat\mu^*,\hat\kappa^*) \le 2\sum_{i=1}^n\{(\hat\kappa^*-\kappa_0)U_i + \hat\alpha^*\hat\mu^*Y_i\} - \sum_{i=1}^n\{(\hat\kappa^*-\kappa_0)^2U_i^2 + \hat\alpha^{*2}\hat\mu^{*2}Y_i^2\}\{1+o_p(1)\}. \tag{28}$$

The leading term of (28) is maximized at $\kappa = \tilde\kappa$ and $\alpha\theta = \widetilde{\alpha\mu}$ with

$$\tilde\kappa - \kappa_0 = \frac{\sum U_i}{\sum U_i^2}, \quad \widetilde{\alpha\mu} = \frac{\sum Y_i}{\sum Y_i^2}. \tag{29}$$

Thus

$$r_{1n}(\hat\alpha^*,\hat\mu^*,\hat\kappa^*) \le \frac{\{\sum U_i\}^2}{\sum U_i^2} + \frac{[\sum Y_i]^2}{\sum Y_i^2} + o_p(1). \tag{30}$$

Let $\tilde{\mu} = \frac{\sum Y_i}{\sum Y_i^2}$ and $\tilde{\alpha} = 1$. Using the sandwich method again, we get

$$r_{1n}(1, \tilde{\mu}, \tilde{\kappa}) = \frac{\{\sum U_i\}^2}{\sum U_i^2} + \frac{\{\sum Y_i\}^2}{\sum Y_i^2} + o_p(1). \tag{31}$$

Combining (31) and (23), we have

$$\frac{\{\sum Y_i\}^2}{\sum Y_i^2} + o_p(1) \leq M_n \leq r_{1n}(1, \tilde{\mu}, \tilde{\kappa}) + r_{2n} = \frac{\{\sum Y_i\}^2}{\sum Y_i^2} + o_p(1).$$

The conclusion is then obvious. □

Figure 1: Q-Q plots of M_n against χ_1^2 for $\kappa = 3$ with $C = 1$.

Figure 2: The comparison of a kernel density estimate (solid line) and the density of the estimated two-component von Mises mixture distribution (dashed line).

Table 1: Simulated null rejection rates (%) of MLRT

		Nominal Levels		
κ	n	10	5	1
1	50	12.37	6.89	1.69
	100	11.55	6.38	1.44
	200	11.12	5.83	1.33
	500	10.90	5.62	1.33
2	50	11.58	6.21	1.61
	100	11.41	6.24	1.68
	200	11.12	5.73	1.28
	500	10.49	5.31	1.09
3	50	11.90	6.80	1.52
	100	10.99	6.14	1.49
	200	10.88	5.92	1.37
	500	10.77	5.41	1.18
4	50	12.03	6.41	1.59
	100	10.89	5.86	1.39
	200	10.84	5.76	1.22
	500	10.68	5.45	1.14